Patent application

of

Hai Shum

for

Title: Method and Apparatus for on Screen Vocabulary

for Chinese Language Word Processing

Background of the Invention

1, Field of the invention

This invention relates to a word processing program for inputting Chinese language into

a modern computer, and more particulary for inputting Chinese language into a computer

with a Chinese language operating system installed.

2, Description of the related prior art

It is surely not a starting revelation to state that the Chinese language is completely

different from the English and other European spelling languages. Typewriting the Chinese

words into a modern computer was not possible until about twenty years ago. Since

then, nearly a thousand inventions to solve this challenging task have patented world

wide. Among these inventions, the majority is the kind that adapted the concept of the English typing; using the alphabet keys of the keyboard to spell out the Latin spelling of the Chinese words; or using the keys, each of which is assigned with a few strokes and a few radicals of the nearly two hundred redicals from the dismentaled Chinese words. Re-build the words while typing the keys. These methods are currently utilized.

For the people, whose first accent is Mandaring or an accent similar to Mandarin. The Latin spelling methods do not require a serious training, used by many non-professional typists occationaly for brief written communications. Due to the fact that many words share similar spelling, to be able to pronounce the words correctly is required before conducting the typing task. Due to the fact that same spelling is usually for a few different words, which are completely different in their meanings as well as in their appearences, distinguished from each other by the four tones; to identify the desired word, to compensate this shortcomming, a list of these words related to the same spelling is shown in a small window, the typist is required to type in a number associates with the word to select it. This selecting process reduces the speed of typing. Often, the typist has to try a few attempts to get the desired word. The Chinese language has nearly one hundred different dialects. For the people, who are not familiar with the standard Mandaring, this spelling method generates huge amount of error, reduces typing speed and accuracy.

The majority of the professional typists use the stroke methods mentioned above.

The main concept of these methods is to disassemble each word into strokes, assigning them to the keyboard; or into the four corners of the word, assigned with codes; or into the radicals assigned with the A to Z alphabets on the keyboard. The typists re-build the words from these components. It is necessary to point out that the English words are

assembled by uniformed alphabets, in a horizontal continuance matter. The Chinese words are two dimentionally assembled by strokes. Furthermore, there are no standard length, no standard angles for the strokes, which is irregularly tailored for each word. Often, one Chinese word can be willingly dismantled into a few different sets of strokes. Although the modern computer and the modern word processing programs are able to recognize the Chinese words, able to re-build the words from the keyboard input of strokes; the typists have to memorize the orders of the assemblies of all the words. After a period of very serious training, the professional typists are able to input Chinese words in high speed. The tough training is fearful. The typing is extensively stressful. It is not a profound strategic solution.

Each Chinese word is a unique piece of art. To keep its integrity has been demanded for generations. It has been proved that the spelling methods are not precise, nor concise. The existing computer keyboard was designed by the West, originally for the English and other European spelling languages. It is a misconception to equalize the alphabets with the strokes. Furthermore, these stroke methods are based on the structure of the word. The Chinese word structures are different between the traditional form and the simplified form. These stroke methods have limitations from their origin. The stroke method is not a permanent solution either. Due to the difficulty of using these word processing methods, people, who is familiar with Chinese language, who is knowledgeable in computer, but still unable to typewrite, unwilling to learn typing is a common phenomenon. Many intellectuals still keep hand writing.

There are some inventions patented to input the integral Chinese words using the cursor to select the words from a list displayed on the screen. This method shared a common

3/4

requirement with the methods described above to have an internal Chinese vocabulary in storage. Instead of match each typed set of Latin spelling or set of strokes/radicals/codes to the corresponding internal storaged word, this method copies the words directly from the word display to the word processor. In addition, this method substantially reduced the dependence of the English keyboard. Among the inventions of this kind, a patent No.CN1034075A (application # 88103689.7) issued by China to Yiping Hou, in July 1989, describes a system and method to input integral words. An U.S. patent No.4951033, issued in Aug. 1990 to Sakaguchi describes nearly the same. However, until the present time, there is no such method available for practical use. These inventions described no details of the arrangement of the Chinese vocabulary in storage, described no details of an efficient word search method. That explained the reason of the vacancy of real usage of the inventions of this kind at the present time. With the assistance of the modern technology, selecting an item from a list displaying on the screen is not a difficult task, is no longer a novelty. An effective, rapid search method of the Chinese vocabulary is a constant, perpetual demand. The Chinese language has thousands words, each word is not an idle item. Each word has meaning, is a dynamic vigorous unit of art.

To list them intelligently, to locate them easily, to select them conveniently, to copy them to the word processor quickly and correctly is a challenging task.

At this stand point, the present invention offers an innovative permanent solution.

Summary

The present invention is aimed to solve the problems mentioned above.

The primary objective of this present invention is to maintain the integrity of the Chinese language. Words are not dismantled into strokes, nor spelled by alphabets. Each word is displayed and entered as a complete integral unit.

The main objective of this present invention is to creat a typewriting method that does not require training. A person with minimum knowledge of computer and capable to read, to write Chinese language is able to typewrite Chinese words using this method offered by this invention, with complete accuracy at the virgin attempt. The pace of typing increases as the typing tasks progresses.

The traditional way to search a word in a Chinese dictionary is following the orders of the orthodox radicals. It is often a very hectic process. Some words have multiple radicals, some words have hidden radicals, some words have implicit radicals and some words have conjectural ambiguous radicals. Some words have multiple meanings.

In a Chinese dictionary, a word appears only once as a principle of a dictionary, despite the extreme difficulty of finding it. A few attempts of trying different radicals is often needed before locating a word or even finding unsuccesfully. This orthodox radical word arrangement principle has restrained the broad use of the traditional Chinese dictionary. Simply adapting it to the vocabulary arrangement for word processing is not wise.

The present invention offers an important objective to arranging the words primarily from the meaning thereof, the radical order is subordinant. Words also filled into pages according to the shape and the dominant stroke thereof. Under this word classifying

principle, some words have opportunity to be listed into multiple pages. A feature is added to have additional blank space in each page for encouraging the user to fill in additional words conveniently for easy locating.

Until the present time, inventions of this class/sub-class, having become practically use are all adapted the concept of English typing, heavily depend on the keyboard. It is another objective of this invention to minimize the dependence of the keyboard. There is another beneficial objective of the invention to be completely independent from the Chinese word pronuciation, users from different region with different dialect/accent background have an equal opportunity to utilize this method. Furthermore, there are no requirement of memorizing any long set of codes.

There is an additional objective of the present invention. The traditional form and the simplified form of Chinese words are stored in two separated divisions, an user has no fear to write a wrong form of word in a formal written communication. However, these two forms of words exist whithin same word processing program offered by this invention. It takes only a click on the pointing device or a tap on the keyboard to switch from one to another. It is convenient for the typist to typewrite an article having these both forms of words in some occations.

It may be worthy to mention that with both forms of Chinese words in storage, the traditional and the simplified, easily switching between them, the user may use it as a referrence for interpretting. It is broadly known that the radicals are the roots of the Chinese words, as mentioned above, words grouped together according to thier logical meaning, thier social or scientific classes, as well as the radicals. This invention offers a good referencing study material for the user to be more familiar with the language.

Brief Description of the Drawings

Fig.1, the flow chart of locating and entering a word.

Fig.2, the home page.

Fig.3, a typical display of a word page.

Fig.4, the complete structure of the Chinese vocabulary storage system.

Fig.5, the flow chart of word relocation.

Detailed Description of the Invention

The present invention splits all available Chinese words into two divisions, the traditional division and the simplified division. The two divisions are identical, in a three dimensional structure, parallelly collocated, consisting three vocabulary tiers. Each tier has 26 pages. Each page is divided into a plurality of addresses, each address occupied by one word. For the reason that the traditional division and the simplified division are symmetrical, each address with its occupant in a division is correspondingly the equivalent to the same address, same occupant in the opposite division.

Fig.1, illustrates a word locating and selecting flow chart, describing the relationship between the tiers, the relationship between the pages and the relationship between a tier with the subordinant pages thereof; describing a typical process of searching, locating and selecting a word. As mentioned above, the two divisions are identical in their structure. One flow chart explains the relationship and the process for both divisions. The step 100 marks the start, introduces the home page 120. Commands are sent from

the interface devices, the keyboard and the pointing device(mouse) 110.

The description of steps 120, the home page, is given in Fig.3.

Unless the user selects other tiers, the program defaults to the primary tier 122, where contains the most frequently used words. Pages are indexed with alphabet letters on the keyboard, as one of the efforts of this invention to use existing equipment. From a to z, use the lower case, for the simplified form Chinese words; use the upper capital case A to Z for the traditional form Chinese words. On the primary and secondary tiers, each page has 100 words, pages on the supplemental tier, each may contain more than 100 words. The primary tier in a division may contain over 2500 words. For a normal daily written communication, this is adequate. The supplemental tier contains the least frequently used words. From other angle of view, each page, indexed by one alphabet letter, gathered words sharing one dominant meaning, includes three tiers, from the most frequently used word tier to the least frequently used word tier.

Steps 131 represents all 26 pages of the most frequently used words, To select between the pages, the user taps the alphabet letter keys on the keyboard with one hand and to select the words using the mouse with another hand. Every word in a page is assigned with an address. For 100 words on one page, a two digit number is enough, one number of the column and one number of the row, is easy to handle. A word is addressed by the page alphabet and the column and row numbers thereof. When this process is familiarized by the user, who may key in the addresses to copy the words to the word processor, no need to display the pages on the screen. If a voice device is equipped, the user may call in the address for the word.

These privileged simple addresses are for the words on the default tier(pages) only.

For the second tier, each address is pre-fixed with the number 2 before the alphabet of the page followed by the column number and the row number. Same rule applied to the supplement tier, but pre-fixed with the number 3. The user may key in these addresses to copy the associated words to the word processor, instead of displaying these tiers/pages on the screen. For example, a word in the primary tier, in a page led by letter "a", may have its address "a47". A less frequently used word in the same word group, classified into the secondary tier, may have its address "2a35". The "a" is for the simplified form, the equivalent traditional form counterparts of these words will occupy the addresses of "A47" and "2A35". A word in the supplemental tier of the same word group may have an address of "3a1209".

Pages within a tier is coplanar. Tiers are for classification only, to arrange words from the most frequently used to the least frequently used. Switching between the tiers is done by using the up/down keys on the keyboard. Tiers are for indexing only. Words are contained in pages. Words under one dominant meaning(more explanation given in Fig.3), or under one other dominant fact (radicals or strokes) are indexed by one alphabet letter. Words under one dominant meaning, or under one other dominant fact are classified into one groups arranged into three tiers as mentioned above. If a desired word is not in the primary page, "page down" to the same group on the second tier, or further "page down" to the same group on the supplemental tier. The three tiers are not coplanar. The step 131 is followed by step 132, if the desired word is located, as in the "Y" (yes) path, the word is copied to the word processor 140. Otherwise, as in the "N" (no) path, the user switch the page to its second tier, as indicated by step 133. If the desired word is located here as in the step 134 (Y) path, it is copied to the word processor 140.

If the word is not located in the second tier (N), switching into the supplemental tier 135. The step 136 is same as step 134, if the word is located. However, if the word is not found, in step 137, this invention gives the user an opportunity to creat the desired word, which is then stored in the page of the tier in step 135.

This completed one word locating and selecting process. After the desired word is copied to the word processor 140, the user may select the next word in the same page, may switch to another page to locate it, step 141 indicates the choice of the user, if no idea of the index of the next word, return to the home page.

Fig.2, illustrates the home page 200, wherein the traditional division 210 and the simplified division 220 are identical in their structure, containing same number of tiers, pages and addresses in exact same location; each word occupying the associated address in one division is the same word equivalent in the exact same address in the opposite division. To find the equivalent word in the opposite division, a user types in the address of the word, changes its capital form with the "Shift + Caps" keys. Or from this home page, selecting the page where the word located, a user returns to the home page to shift to the opposite division to find the equivalent word. The traditional form is used in Taiwan, Hong Kong and most oversea areas, the simplified form is used in the mainland China, Singapore and some other Asian countries. Due to a long period of isolation, people has difficulty to recognize words in the opposite form. Dictionaries usually list all the available words undiscriminatory, commingle words of these two forms. Other Chinese word processing programs are having all Chinese words mixed. From a common knowledge, for communicating formally in written language, to keep all words in an unified form is a good manner. The present invention offers the separation of these two forms. But finding

the equivalent word is easy with this cross reference tool.

Within the areas of 210 or 220, three choices: default, second, supplement are listed for direct access to these three tiers. Without making the selections, the program automatically is in the default seting for the primary tier of vocabulary. Area 230 is a matrix of 27 keys for the index of the word pages, offering direct access to the word pages by pointing the cursor and click on them. Again, the default seting is for the primary words. It changes to other two tiers when selected. The 26 keys are indexed by the English alphabet letters on the keyboard, leaving the last one blank for additional expansion. To index the Chinese words for listing intelligently, for searching easily, is always a very challenging task for centuries. Until the present time, there is no perminant best solution for every one to accept it. A principle of this invention is to utilize the existing equipment of a common personal computer, to use the 26 English letters for the indexing seems to be convenient. To classify the thousands of Chinese words into 26 groups, the best way seems to be gathering the words by their dominant meanings. The present invention classifies the word meanings into 23 groups led by letters from A to W, uses X,Y,Z for words difficult to classify, for special words and some very basic words. The words are arranged into these 23 groups by their dominant meanings of most basic elements of the universe, such as sun, moon, sky; earth, rock; domestic mammal; fish, water, boat; human, male, female; metal, tool, knife; measurement system; wood, bamboo.... A person needs only a basic knowledge of these elements to directly lock into the related page to locate the desired Chinese words. Words under one dominant meaning are gathered in one group, indexed by one letter, divided to three tiers. The primary tier has 100 most frequently used words, in most occations, these words are sufficient. This primary

tier is the default seting. The second tier has 100 less important words. The supplemental

tier contains all words that are not listed in the other two higher priority tiers. This tier

has no limit for listing. New words are added or created. Words may be relocated from

other two tiers to this tier, or move out of. For the reason of no limit for word listing,

the individual address in this tier is no limit to a two digit number. A modern Chinese

word processor, such as the Microsoft Windows 98, has the ability to creat new words.

This invention is intended to be a feature of a Chinese word processor. A modern

Chinese word processor further offers a function of associated compound word,

immediately after a word in entered. The punctuation marks are arranged as well.

Area 240 is a key leading to a detailed listing of all the pages. A new user will find

it very handy to look up the indexes with all the "dominant meanings" they carriy. It is

shown in Fig.1 as step 121. Area 250 is a key leading to the word relocation process,

detail given in Fig.5.

Fig.3 is a typical word processor display 300, with an inset window 310. Showing here

is a page in the default tier of 100 primary words under the dominant meaning of

domestic mammals. Bar 311 is for a cursor to be landed and drag the inset window 310

to any location to yield the main window for word processing. The inset window 310 is

floating automatically as the typing in progress. Four keys 312 are for the purpose of the

following: (h) key for returning to home page, (+) key for enlargement of the inset window

310, (-) kay for reduction of it, (x) key for cancelling it.

Area 320 is the matrix of the vocabulary. The top row 321 is the ten Chinese numerals

from 1 to 0. The left column 322 is the ten Arabic numerals from 1 to 0. Both of these

numerals can be selected for word processing, as well as for indication of the column

and row number for addressing the words, but are not assigned with addresses therefor.

The dominant meaning of this page is domestic mammals, includes: sheep, dog, pig, horse, cow, cat , rat, elephant and camel. Words listed here are either direct relatives of these mammals, such as "fox" and "wolf" are listed with "dog"; or words derived from these mammals, such as "drive" is listed with "horse", "plough" is with "cow". Due to the page limitation, "tiger" is not listed here with "cat". There is another group for wild mammals. Interestingly, "monky" is here for the reason that it has the "dog" radical. In another page of "water, fish and boat", one will find that "whale" is listed for the same reason of having the "fish" radical, the ancient Chinese created this word long before the modern bio-science discovered that whale was not a fish. But whale lives with fish in water.

The top row of Chinese words in the window, ten of them in row 1, are all pronounced "Yang". There are more words with this same pronunciation. None of them shares similar meaning with others. An example of how complicated this language is.

The word occupies the address 11 (the one at the top left corner) is "oxygen". Listed here for the reason of having same pronuciation of "sheep", having the radical of "sheep". For the reason of having a co-radical of "air", it will be listed again in the page of "air, weather, atmosphere". For the reason of being a frequently used word, it will be listed in another primary page as well. A language is a major part of a culture. It may not be logical. A user of this program, knowing Chinese language, will find no problem to locate words. Again, some words may be listed in more than one groups.

The alphabet letter 323 at the corner is the page index. Clicking at this letter to switch to the other two pages in this word group, to find the less frequently used words, or to return to the default page. As already mentioned, this letter is a part of the addresses

for the words on this page. The alphabet letter keys of the keyboard are no longer for spelling the English words, instead, given a new use of indexing the word pages and for addressing the words thereof. An important feature is added to this program letting the user to relocate words. Detail is given in Fig.5.

Fig.4 illustrates the complete structure of the vocabulary arrengement. Block 400 indicates the Chinese words splited to two divisions: the traditional division 410 and the simplified division 460. Again, these two divisions are identical from the entire structure to every individual address setting. Each pair of the two equivalent words of the two forms is located in the corresponding addresses of the two divisions. This arrangement gives the user a very easy way to find the corresponding word in the opposite form.

The block 470 is the default tier, it is automatically selected for locating primary words. The block 471 shows a to z pages. The block 480 and 481 are second tier and pages, contains same number of pages as the primary tier. The block 490 and 491 are supplement tier and pages, contains no limit of words.

These tiers are not coplanar. A bridge 478 crosses from block 471 to block 481 indicates that words may be relocated between them. A bridge 489 crosses from block 481 to block 491 indicates that words may be relocated between them. A bridge 479 crosses from block 471 to block 491 indicates that words may be relocated between them.

The simplified division components are indexed with lower case of letters a to z.

For the reason of being identical to each other, the traditional division has same number of blocks and bridges as the simplified division. The default tier and pages are 420 and 421. The second tier and pages are 430 and 431. The supplement tier and pages are 440 and 441. The bridges cross them are 423, 434, 424.

For cross reference of the vocabulary of the two word form divisions, bridges are arranged. For the reason of corresponding location of the equivalent words, bridges across the corresponding addresses in the corresponding pages only. That is a bridge across between two same alphabet letter indexed pages only (one page indexed with the upper case letter, the other page with the lower case letter) in a same level of tier. On Fig.4, bridge 427 crosses the pages 421 and 471 for the tiers 420 and 470; bridge 438 crosses the pages 431 and 481 for the tiers 430 and 480; bridge 449 crosses the pages 441 and 491 for the tiers 440 and 490. Again, for the reason of being correspondingly equivalent of the addresses and its occupants(words) in these two divisions; if a word in one division is relocated from one tier to another, the equivalent word in the other division will be relocated automatically to comply with this principle. To emphasize this three dimentional vocabulary arrangement, a x-y-z coordinate 450 is added at the corner. It is for an indication only. It has no value to the structure, is not a part of the invention.

Fig.5 presents a routing of word relocation. At the starting 500, the home page 510 is shown on computer screen. The user click on the key 250 on the home page(Fig.2) for selecting the "word relocate" 520, to display the old addresses window 530 onto the screen. Three address blanks 531 are available for up to three corresponding words to be relocated simultaneously. For words in the primary tiers, an address contains one alphabet letter followed by one digit column number and one digit row number; in the secondary tiers, an address contains one digit prefix number 2 followed by one letter, one digit column number and one digit row number; in the supplemental tiers, an address contains one digit prefix number 3 followed by one letter, column number(s) and row

number(s). Each supplemental tier word page may have more than 100 words, needs two digit column number and two digit row number. Each address blank 531 has six digit space to be fully adequate for accommodation of a full length address.

The user enters the associating address of the word into the blank 531, hit Enter key of computer keyboard, the word will be delievered for identification. Choices are provided on step 540, if the word address is entered incorrectly, return to step 530 to repeat the process. If the word is the desired for relocation, as delieverd to the inset window 551 of new address window 550, the user proceeds to enter the new address(new location) of the word into blank 552. This relocation process is done 560, when Enter key is hit, the word and the address blinked.